© 1994 Oxford University Press

Nucleic Acids Research, 1994, Vol. 22, No. 20 4125–4131

Characterization of the minimal DNA-binding domain of the HIV integrase protein

Ramon A.Puras Lutzke, Cornelis Vink and Ronald H.A.Plasterk*
Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received July 6, 1994; Revised and Accepted September 2, 1994

ABSTRACT

The human immunodeficiency virus (HIV) integrase (IN) protein mediates an essential step in the retroviral lifecycle, the integration of viral DNA into human DNA. A DNA-binding domain of HIV IN has previously been identified in the C-terminal part of the protein. We tested truncated proteins of the C-terminal region of HIV-1 IN for DNA binding activity in two different assays: UV-crosslinking and southwestern blot analysis. We found that a polypeptide fragment of 50 amino acids ($IN_{220-270}$) is sufficient for DNA binding. In contrast to full-length IN protein, this domain is soluble under low salt conditions. DNA binding of $IN_{220-270}$ to both viral DNA and non-specific DNA occurs in an ion-independent fashion. Point mutations were introduced in 10 different amino acid residues of the DNA-binding domain of HIV-2 IN. Mutation of basic amino acid K264 results in strong reduction of DNA binding and of integrase activity.

INTRODUCTION

Retroviral DNA integration into the human genome is an essential step in HIV replication (1,2). After reverse transcription of HIV genomic RNA, the viral integrase (IN) protein mediates the DNA integration process by catalyzing two distinct reactions: (i) site-specific cleavage of two nucleotides from both 3′ viral DNA ends (donor cut) and (ii) integration of the recessed viral DNA into the host genome (for recent reviews see ref. (3–5)). Purified, recombinant HIV IN was found to mediate donor cleavage and strand transfer (6–12) as well as disintegration, which was described as the reversal of the integration reaction (13). By analysis of deletion mutants of IN, three independent domains have been identified in the protein (14,15). The N-terminal region contains a zinc-finger motif (His-X_3-His-X_{20-30}-Cys-X_2-Cys), which is conserved among retroviral and retrotransposon IN proteins (16,17). Although the function of this domain is not clear, it was found that this region can bind zinc ions (14,18) and it is necessary for specific IN activities (14,15). The second conserved region is the central catalytic domain. It contains the conserved DD(35)E motif (16,17) with the putative active site residues D64, D116 and E152 (2,11,19–21). In the C-terminal part of the IN protein a DNA-binding region was identified (15,16,22,23). The DNA-binding characteristics of the full-length protein and deletion derivatives which contain the C-terminal region are similar in the way that they bind both viral DNA and non-specific DNA with approximately similar efficiencies (10,22–25). From kinetic studies of IN from Rous sarcoma virus (RSV), it was suggested that IN is active as dimer or multimer (26). Protein mixing experiments, in which HIV IN proteins mutated in different domains could complement each other (27,28), also demonstrated an active multimeric state of IN (for discussion see Vink and Plasterk (3)).

In order to further characterize the DNA-binding region of IN, we defined the boundaries of this domain by deletion analysis and studied its biochemical properties. The minimal DNA-binding domain of HIV-1 IN was expressed in *Escherichia coli* and purified to homogeneity. In addition, we investigated the role of several (partially) conserved amino acid residues in this region.

MATERIALS AND METHODS

DNA techniques

DNA sequences encoding polypeptides of the C-terminal part of HIV-1 IN were fused to the *malE* gene in expression vector pMAL-c (New England Biolabs). To clone these DNA fragments, a polymerase chain reaction (PCR) was performed using oligonucleotides containing either a *Bam*HI, *Nhe*I or *Spe*I restriction site. The IN gene in plasmid pRP619 (15) was used as template DNA in the PCR reaction. The *Bam*HI/*Spe*I or *Bam*H1/*Nhe*1 PCR-fragments were purified and ligated to the *Bam*HI/*Xba*I digested vector pMAL-c.

To clone the DNA fragment of HIV-1 IN encoding the minimal DNA-binding region ($IN_{220-270}$) a PCR reaction was performed using pRP619 as template and the following oligonucleotides as primer:

AB4039: AGGTCACATATGATTCAAAATTTTCGGGTT(*Nde*I)
and AB4041: GCCATCGGATCCTTAGTGGTGATGGTGATGGTGATCC-CTAATGATCTT (*Bam*HI).

The latter oligonucleotide contains six codons for a 6×histidine tag (6×His tag). The amplified DNA fragment was gel purified, digested with *Bam*HI and *Nde*I and subsequently cloned into the expression vector pET3c (29) resulting in plasmid pRP941.

*To whom correspondence should be addressed

Point mutations in HIV-2 IN (see Fig. 5C) were generated by site-directed mutagenesis as described by van Gent *et al.* (21). DNA fragments encoding the DNA binding region of HIV-2 IN (amino acids 210–270) and containing point mutations were amplified in a PCR reaction with the following oligonucleotides:

AB3839: AGGTCAGGATCCATACAATTCCTCCAAGCCAAA (*Bam*HI) and AB3840: AGTGCCACTAGTCTCTGATGATCTTGGC (*Spe*I).

After purification, these 192 bp fragments were cloned into *Bam*HI/*Xba*I sites of pMALc.

Expression and purification of proteins

E.coli strain JM101 was transformed with the various pMal-c derivatives. Upon induction with 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG), the IN polypeptides fused to maltose binding protein (MBP) were purified by an amylose affinity column as described previously (15).

Wild type and mutant HIV-2 IN proteins were expressed in *E.coli* strain BL21(DE3) pLysS, purified and immobilized on thiopropyl Sepharose beads as described before (21).

Plasmid pRP941 was introduced into BL21(DE3) pLysS and expression was induced by adding 0.3 mM IPTG. Cells were lysed under denaturing conditions in buffer H (25 mM Hepes pH 7.5, 1 M NaCl, 6 M guanidiniumchloride, 0.1 mM EDTA and 1 mM β-mercaptoethanol) supplemented with 5 mM imidazole. After sonication, the lysate was cleared by centrifugation (10,000×g for 20 min) and diluted with one volume buffer H (without NaCl). The cleared lysate was loaded on a nickel-chelating column (Ni-NTA; Qiagen) and washed twice with buffer H (5 and 20 mM imidazole, respectively). Protein $IN_{220-270}$6×His was eluted in buffer H (200 mM imidazole) and peak fractions were further purified by gelfiltration over a Sephacryl S-100 (Pharmacia) column adjusted in buffer H containing 0.5 M NaCl. The eluted protein was renatured by dialysis against 25 mM Hepes pH 7.5, 100 mM NaCl, 0.1 mM EDTA and 1 mM β-mercaptoethanol at 4 °C. Protein was subsequently loaded on a Heparin Sepharose CL-6B (Pharmacia) column and eluted with a linear gradient from 0.1 to 2 M NaCl in buffer H (without guanidiniumchloride). Thehomogeneity of $IN_{220-270}$6×His was analyzed on denaturing 16.5% tricine polyacrylamide gels (30).

DNA binding assays

Southwestern blot analysis was carried out as described by Woerner *et al.* (23) using plasmid PiAN7 as (non-specific) probe. The plasmid was radiolabeled by random priming in the presence of (α-^{32}P)dATP (31).

In UV-crosslinking experiments, typically 20 nM ^{32}P-labelled viral oligonucleotides (TTAGTCAGTGTGGAAAATCTCTA-GCAGT) annealed with the complement bottom strand, 0.3 μM protein (MBP- or 6×His-fusion protein) 25 mM Hepes pH 7.5, 50 mM NaCl, 1 mM $MnCl_2$ (or when indicated 1 mM $MgCl_2$, without divalent cations, 1 or 2 mM EDTA, respectively), 0.5% Tween 20, 2% glycerol and 0.1 mM β-mercaptoethanol were incubated in a volume of 20 μl for 30 min on ice. Reaction mixtures were irradiated for 5 min on ice on a 'Chromato-vue' transluminator (254 nm, 50 Hz, 0.6 A, Ultraviolet products Inc.) and stopped by adding one volume protein loading buffer.

Cleavage assay

Site-specific cleavage of oligonucleotides containing HIV-2 U5 LTR sequences was performed as described by van Gent *et al.* (21). Reactions were analyzed on a 12% denaturing polyacrylamide gel followed by autoradiography. Activities were quantified by densitometry, using an Ultrascan XL Enhanced Laser Densitometer (LKB).

RESULTS

Fine-mapping of the DNA binding domain of HIV-1 IN

To determine the minimal region of HIV-1 integrase necessary for DNA binding, we generated a set of DNA constructs which encode various parts of the C-terminus of HIV-1 IN. The DNA fragments were fsed to the 3′ end of the *malE* gene in vector pMal-c. The various polypeptides were expressed as fusions to maltose-binding protein (MBP), and purified by affinity chromatography. In figure 1D the MBP–IN fusion proteins are listed. Purified fusion proteins were analyzed by SDS–polyacrylamide electrophoresis (SDS–PAGE) as shown in Fig. 1A. We assayed these proteins for DNA binding activity by Southwestern blot analysis, using a non-specific DNA probe (Fig.1B). Truncated IN polypeptides containing amino acids 225 to 270 (lanes 3–7) possess wild type IN DNA-binding activity. Polypeptide MBP–$IN_{200-265}$ (lane 11) has a weaker activity than the wild type IN protein (MBP–WT; lane 2). For MBP–$IN_{230-270}$ (lane 8) only a marginal DNA-binding activity could be detected (after prolonged exposure; data not shown). Fusion proteins MBP–$IN_{240-270}$ and MBP–$IN_{230-260}$ (lanes 9 and 10) were not capable of binding DNA.

We also tested DNA-binding activity by protein–DNA UV-crosslinking experiments. Proteins were incubated with radiolabeled DNA substrate, irradiated with UV-light (see Materials and Methods) and analyzed by SDS–PAGE followed by autoradiography (Fig.1C). The estimated molecular weight of the major protein–DNA complex is approximately 80,000, which could correspond to a MBP–IN WT protein monomer (approximately 74 kDa) bound to a single-stranded 28-nucleotide DNA substrate (approximately 9 kDa). Also minor UV-crosslinked species could be observed with a lower mobility than the major UV-crosslinked product. The composition of these complexes is not understood; the molecular weight of these products is not in accordance with multimeric forms of IN bound to DNA. Proteins containing amino acids 220–270 (lanes 5–8) bind DNA with a similar efficiency as WT IN, while proteins containing IN sequences 225–270 and 230–270 (lanes 9 and 10) show a strongly reduced DNA-binding activity. MBP–$IN_{200-265}$, MBP–$IN_{240-270}$ and MBP–$IN_{230-260}$ were not able to bind DNA.

The results from southwestern blot analysis and UV-crosslinking analysis are summarized in figure 1D. Although there is a slight difference in DNA-binding activities in the two assays, our results demonstrate that a region of 50 amino acids, located between IN amino acids 220 and 270, is sufficient for DNA binding.

To determine whether DNA-binding activities obtained with the MBP-fusion proteins were influenced by the MBP part of the fusion protein, we tested a polypeptide containing amino acids 220–270 of HIV-1 IN as fusion protein to a C-terminal 6×His tag. The corresponding protein of approximately 6 kDa was purified under denaturing conditions (see Materials and Methods) and the homogeneity of the protein preparation was analyzed by SDS–PAGE (Fig.2). In contrast to the full-length IN protein, $IN_{220-270}$6×His is soluble in 50 mM NaCl after renaturation

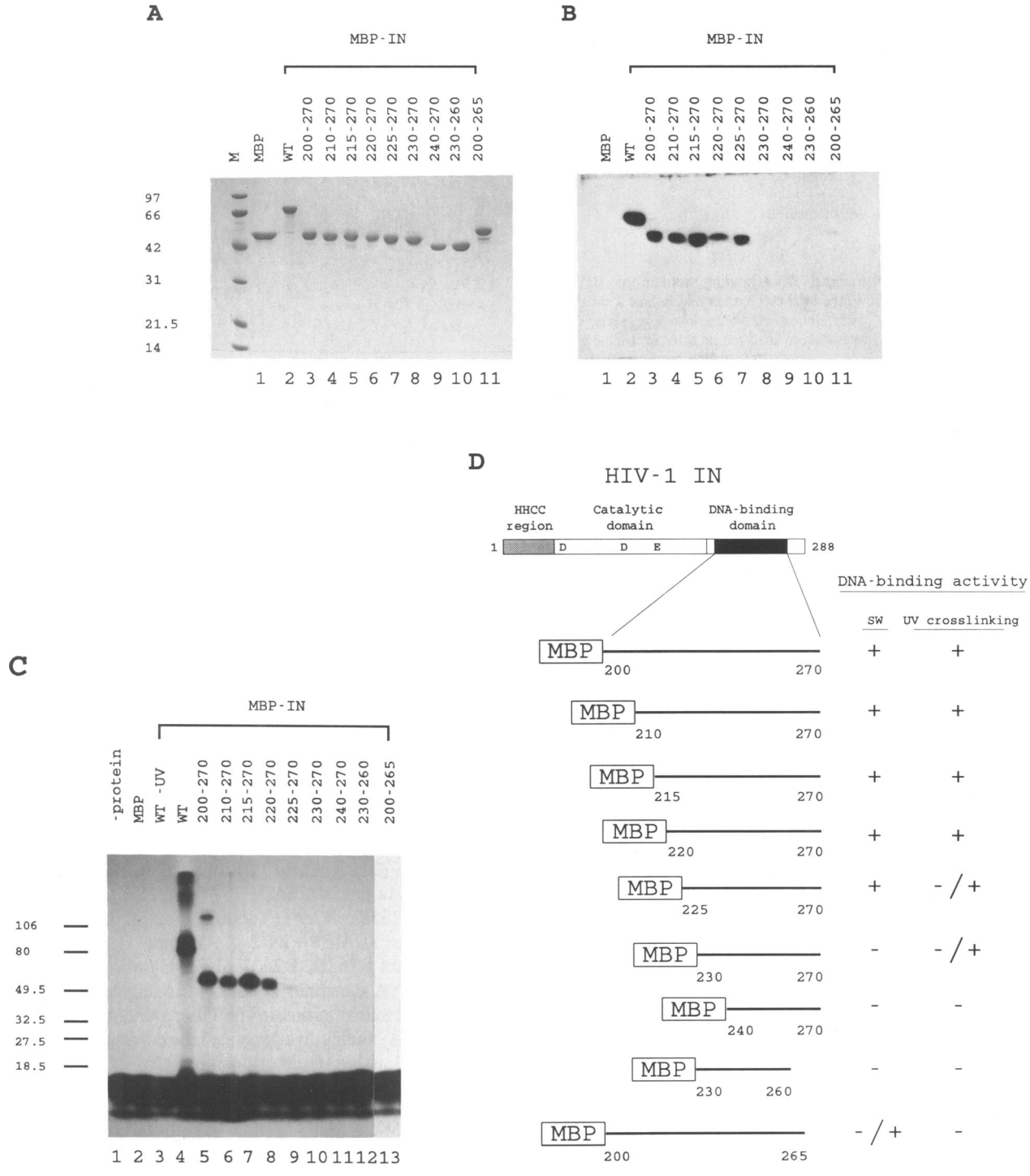

Figure 1. DNA-binding analysis of the C-terminal IN domain. (**A**) Truncated IN polypeptides fused to maltose binding protein (MBP) were analyzed on a 10% SDS−PAGE, followed by coomassie staining. Molecular weight markers (M) are indicated at the left in kilodaltons. (**B**) Southwestern blot analysis was carried out as described (23) with the same proteins as in (A) and radiolabeled plasmid PiAN7 was used as non-specific probe. As negative control MBP−β-galα (MBP; lane 1) was tested. DNA binding of MBP−IN proteins was detected by autoradiography. (**C**) UV-crosslinking analysis. Covalent protein−DNA complexes were analyzed on a 10% PAGE. Molecular weight markers are indicated at the left in kilodaltons. In lane 1 the free substrate is shown (−protein), in lane 2 MBP−β-galα (MBP). Lane 3 shows a reaction with MBP−IN WT in which UV irradiation was omitted (WT −UV). (**D**) Schematic presentation of all MBP−IN fusion proteins and their DNA-binding activities, analyzed by Southwestern blot analysis (sw) and UV-crosslinking.

and does not seem to aggregate when it is concentrated up to 20 mg/ml (data not shown). Similar to the MBP-fusion proteins, the $IN_{220-270}6\times$His protein binds DNA in both DNA-binding assays southwestern blot analysis (Fig.2 lane 3) and UV-crosslinking (Fig.3B). Similarly as WT IN, $IN_{220-270}6\times$His did not show any specificity in DNA binding (data not shown).

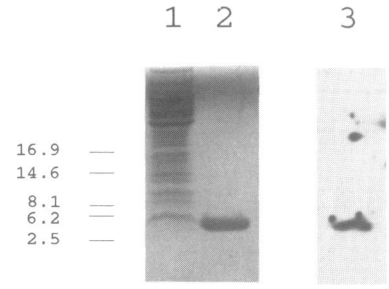

Figure 2. Analysis of the minimal DNA-binding domain of HIV-1 IN ($IN_{220-270}6\times His$). Lane 1 shows the total cell extract and in lane 2 the purified polypeptide $IN_{220-270}6\times His$, after elution from a heparin−sepharose column (see Materials and Methods). Proteins were analyzed on a 16.5% tricine gel (30), followed by coomassie staining (a low molecular weight marker is indicated at the left in kilodaltons). DNA binding of $IN_{220-270}6\times His$ was detected by Southwestern blot analysis (lane 3) using radiolabeled PiAN7 as probe.

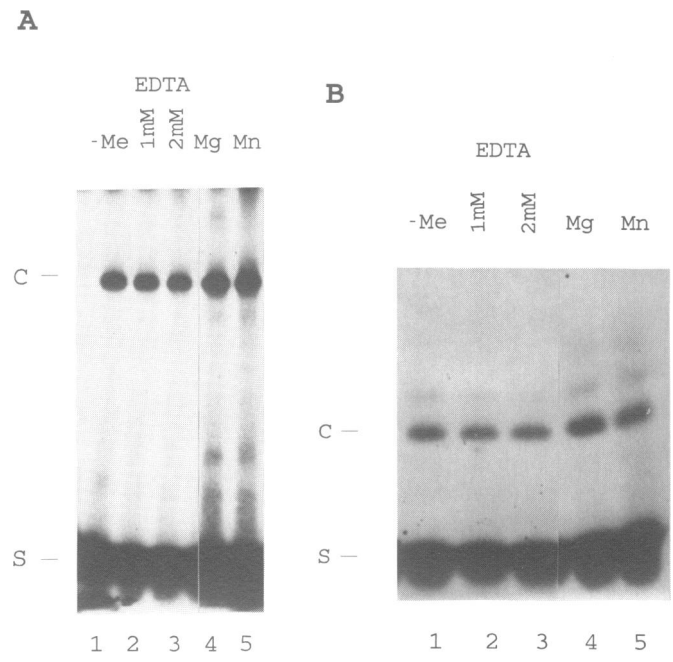

Figure 3. Requirement for divalent cations in DNA binding analyzed by UV crosslinking. (**A**) Effects of Mn^{2+} (lane 5) and Mg^{2+} (lane 4) on crosslinking of MBP−$IN_{215-270}$ to the viral DNA substrate (see Fig.1C). In lanes 1 (−Me), 2 and 3 divalent cations were omitted and in lanes 2 and 3 reactions were carried out in the presence of EDTA (1 and 2 mM, respectively). (**B**) Similar UV-crosslinking experiments as in (B) but with the His tagged polypeptide $IN_{220-270}6\times His$. The radiolabeled DNA substrate is indicated by S and the UV crosslinked protein−DNA complex by C.

Ion-dependency of the DNA-binding domain of HIV-1 IN

Recombinant, purified HIV-1 IN requires divalent cations for its catalytic activities (6,8). To investigate whether the DNA binding domain of IN requires divalent cations for DNA-binding activity, we performed UV-crosslinking experiments in the presence of either Mn^{2+}or Mg^{2+} (Fig.3). Both the MBP-fusion protein (MBP−$IN_{215-270}$; Fig.3A) and the His-tagged protein ($IN_{220-270}6\times His$; Fig.3B) bind DNA either in the absence or presence of divalent cations. The UV-crosslinking efficiency is somewhat more efficient in the presence of either Mg^{2+} or Mn^{2+}. We also performed UV-crosslinking analysis in the presence of EDTA to remove traces of divalent cations, which could remain associated with the protein during protein purification (Fig.3 lanes 2 and 3). The level of DNA binding was only slightly reduced in the presence of EDTA than in the presence of divalent cations. This indicates that the integrase DNA-binding domain ($IN_{220-270}$) binds DNA in an ion-independent fashion.

Mutational analysis of DNA-binding domain of HIV-2 IN

The DNA-binding domain of IN does not share obvious similarities with any of the known DNA-binding motifs, such as helix-turn-helix or zinc-finger motifs. Characteristic for this domain is the prevailing presence of basic amino acids (Fig.4); polypeptide $IN_{220-270}$ has a predicted isoelectric point of 11.2. Because this domain only possesses non-specific DNA-binding activity, it is likely that the basic amino acids are involved in DNA binding via the phosphate backbone of the DNA substrate, rather than specific recognition of DNA bases. From amino acid sequence alignment of several retroviral integrase proteins, no highly conserved amino acids could be identified (Fig.4), except for tryptophan at position 235 of HIV-1 IN, which is conserved among all integrases. In addition, several less well conserved amino acid residues can be seen. In order to identify amino acids which could be involved in protein−DNA contacts, we introduced point mutations in the HIV-2 IN gene and purified the corresponding mutant proteins. We selected mainly basic and aromatic amino acids and changed them to either polar or non-polar amino acids, changing two or three amino acid positions at once, as depicted in figure 5C.

First, we tested the DNA-binding ability of the mutants in the context of the DNA-binding domain of HIV-2 IN (MBP−IN-$2_{210-270}$) by UV-crosslinking (Fig.5A). MBP−IN-$2_{210-270}$ binds DNA (lane 1). Also MBP-fusion proteins containing amino acid substitutions Y226A, F227Y, R228E (triple mutant, lane 2) and W235K, K236D (double mutant, lane 3) bind DNA equally well as MBP−IN-$2_{210-270}$ (lane 1). However, another triple mutant R262D, R263V, K264E (lane 4) was not able to bind the DNA substrate. To further investigate which of the three amino acid substitutions in the latter mutant is responsible for loss of DNA binding, we generated the corresponding single point mutants. Mutation of either position R262 or R263 (lanes 5 and 7, respectively) does not result in altered in DNA binding activity compared to WT IN. However, mutation K264E shows strongly reduced DNA binding activity (lane 6).

We also investigated the effect of these mutations in the context of the full-length HIV-2 IN protein on specific cleavage and integration activity. As shown in Fig.5B the triple mutant Y226A, F227G, R228E (lane 2), double mutant W235K, K236D (lane 3) and single point mutant R263V (lanes 7) process the viral DNA termini to the same extent as WT HIV-2 IN protein (lane 1). Reduced cleavage activities were observed with point mutants R262D (42% of WT activity; lane 5) and K264E (7% of WT activity; lane 6). The DNA-binding deficient triple mutant R262D, R263V, K264E is not able to cleave the viral DNA substrate (lane 4). The integration activities of the mutants were comparable to the corresponding cleavage activities in Fig. 5C (data not shown).

```
           220       230           240       250       260       270
           ↓         ↓             ↓         ↓         ↓         ↓
HIV-1      IQNFRVYYRDSRN....PLWKGPAKLLWKGEGAVVIQD..NSDIKVVPRRKAKIIRD
HIV-2      LKDFRVYFREGRD....QLWKGPGELLWKGEGAVLVKV..GTDIKIIPRRKAKIIRD
FIV        LQAQWIYYKDQKD....KKWKGPMRVEYWGQGSVLLKD.EEKGYFLIPRRHIRRVPE
SIV        FKNFRVYYREGRD....QLWKGPGELLWKGEGAVILKV..GTDIKVVPRRKAKIIKD
BIV        KIEKWCYVRNRR.....KEWKGPYKVLWDGDGAAVIEE..EGKTALYPHRHMRFIPP
E1AV       SSKKFCFYKIPGE....HDWKGPTRVLWKGDGAVVVND.EGKGIIAVPLTRTKLLIK
VISNA      EKIRFCYYRTRKR.GHPGEWQGPTQVLWGGDGAIVVKDRGTDRYLVIANKDVKFIPP
MPMV       KQFAMVKWKDPLD....NTWHGPDPVLIWGRGSVCVYSQTYDAARWLPERLVRQVSN
RSV        TEGPPVKIRIET.....GEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDIT
MMTV       DPKPMVMWKDLLT....GSWKGPDVLITAGRGYACVFPQDAETPIWVPDRFIRPFTE
HTLV-I     KQTHWYYFKLPGL..NSRQWKGPQEALQEAAGAALIPV.SASSAQWIPWRLLKRAAC
HTLV-II    PPPKWFYYKLPGL..TNQRWKGPLQSLQEAAGAALLSI..DGSPRWIPWRFLKKAAC
BLV        GSDKLFLYLLPGQ..NNRRWLGPLPALVEASGGALLAT...DPPVWVPWRLLKAFKC
MoMLV      RIGDSVWVRRHQTKNLEPRWKGPYTVLLTTPTALKVDG....ISAWIHAAHVKAATT

                              *
consensus  ......YY..........WKGP..VLW.G.GA..V.........W.P.R..K....
```

Figure 4. Alignment of retroviral integrases. HIV-1 IN (41,42) is aligned with other retroviral integrases. The non-specific DNA-binding domain ($IN_{220-270}$) is compared with corresponding amino acid sequences of other retroviruses. Amino acid positions of HIV-1 $IN_{220-270}$ are indicated by arrows. HIV-2, human immunodeficiency virus type 2 (43); FIV, feline immunodeficiency virus (44); SIV, simian immunodeficiency virus (45); BIV, bovine immunodeficiency virus (46); E1AV, equine infectious anemia virus (47,48); VISNA, visna virus (49); MPMV, Mason–Pfizer monkey virus (50); RSV, Rous sarcoma virus (51); MMTV, mouse mammary tumor virus (52); HTLV-I and HTLV-II, human T-cell leukemia virus type I (53) and II (54); BLV, bovine leukemia virus (55); MoMLV, mouse moloney leukemia virus (56). Positions with >50% identity are depicted with capital letters and position with 100% identity is indicated with an asterisk.

In summary, we identified a stretch of three amino acids R262, R263 and K264 in the DNA-binding domain of HIV-2 IN which is necessary for DNA binding activity, site-specific cleavage and integration activity. Mutation K264E contributed most strongly to the reduced activity seen in the corresponding triple mutant, suggesting that K264 plays a criticalrole in DNA binding. Furthermore, all mutants tested were able to perform disintegration (data not shown), indicating that loss of IN activities is due to diminished DNA-binding activity and not due to decreased catalytic activity.

DISCUSSION

The viral integrase protein contains three functional domains: The N-terminal region, the catalytic domain and a DNA-binding domain (14–16,22,23). Previously, DNA-binding activity of IN has been mapped to the C-terminus of the protein (15,16,22,23). Woerner and Marcus-Sekura (32) located the DNA-binding domain of HIV-1 IN between amino acids 180 and 248. However, their study did not include a characterization of an independently expressed and purified domain which contain amino acids 180 to 248. Vink *et al.* (15) identified a non-specific DNA-binding domain of HIV-1 IN between amino acids 200–270; polypeptide $IN_{200-270}$ fused to maltose-binding protein (MBP) was able to bind DNA. To further define the exact boundaries of the DNA-binding domain, we generated a set of truncated IN polypeptides fused to MBP. Fusion proteins of IN deletion derivatives can be purified in a similar fashion. Comparing the results from UV-crosslinking experiments and southwestern blot analysis, we conclude that the non-specific DNA-binding domain of HIV-1 IN is located between amino acids 220 and 270 (see Fig. 1D). Polypeptide $IN_{220-270}6\times$His is also able to bind DNA in both DNA-binding assays, supporting the notion that MBP does not have any influence on DNA-binding activities of the fusion proteins.

Although there are hardly any conserved amino acid residues in the DNA-binding domain of IN and other related retroviral integrases, basic amino acids are predominant in this region of IN (see Fig.4). This main composition of basic amino acids indicates that they could be involved in non-specific DNA binding via contacts with the phosphate backbone ofthe DNA. To analyze possible amino acid candidates for IN–DNA interaction, we made a set of point mutants by site-directed mutagenesis and tested their DNA-binding ability (see Fig.5C). We found a stretch of three amino acids in HIV-2 IN (R262, R263, K264) which together play a critical role in DNA binding. Further analysis of these three amino acids showed that changing lysine at position 264 strongly reduces DNA binding. This suggests that this residue is possibly involved in IN–DNA interaction. Surprisingly, mutation in the only conserved amino acid in the DNA-binding domain of HIV-2 IN, tryptophan 235, does not lead to reduced DNA-binding activity and cleavage activity. Similarly, mutant W235E of HIV-1 IN was shown to have wild type activity in cleavage and integration assays (33). However, little is known about mutations in the C-terminal part of IN and their effects on retroviral replication *in vivo*. It would be interesting to introduce the mutations that we describe here, in a proviral clone and analyze the activity of the mutant proteins.

Recombinant HIV integrase needs divalent cations for catalytic activity (6,8). It has been suggested, that IN coordinates divalent cations in the active site (19,20), similar to DNA polymerase I of *E.coli* (34). We investigated whether DNA binding by the non-specific DNA binding domain $IN_{220-270}$ is ion dependent. Our results indicate that DNA binding of $IN_{220-270}$ is not ion dependent and does not prefer Mn^{2+} over Mg^{2+} as observed for the site-specific cleavage and integration activity of the full-length protein (6,8).

IN discriminates between specific, viral DNA and non-specific DNA in functional in vitro assays (7,9,24,35). How does IN specifically recognize the viral DNA termini? We and others (10,22,23,25,36) could not demonstrate specific DNA binding of IN to viral LTR sequences. It has been proposed by Vincent *et al.* (12), that the N-terminal HHCC region is involved in specific recognition of viral DNA termini. Here we defined a non specific, ion-independent DNA binding domain of IN, but we cannot rule out the existence of another DNA-binding region,

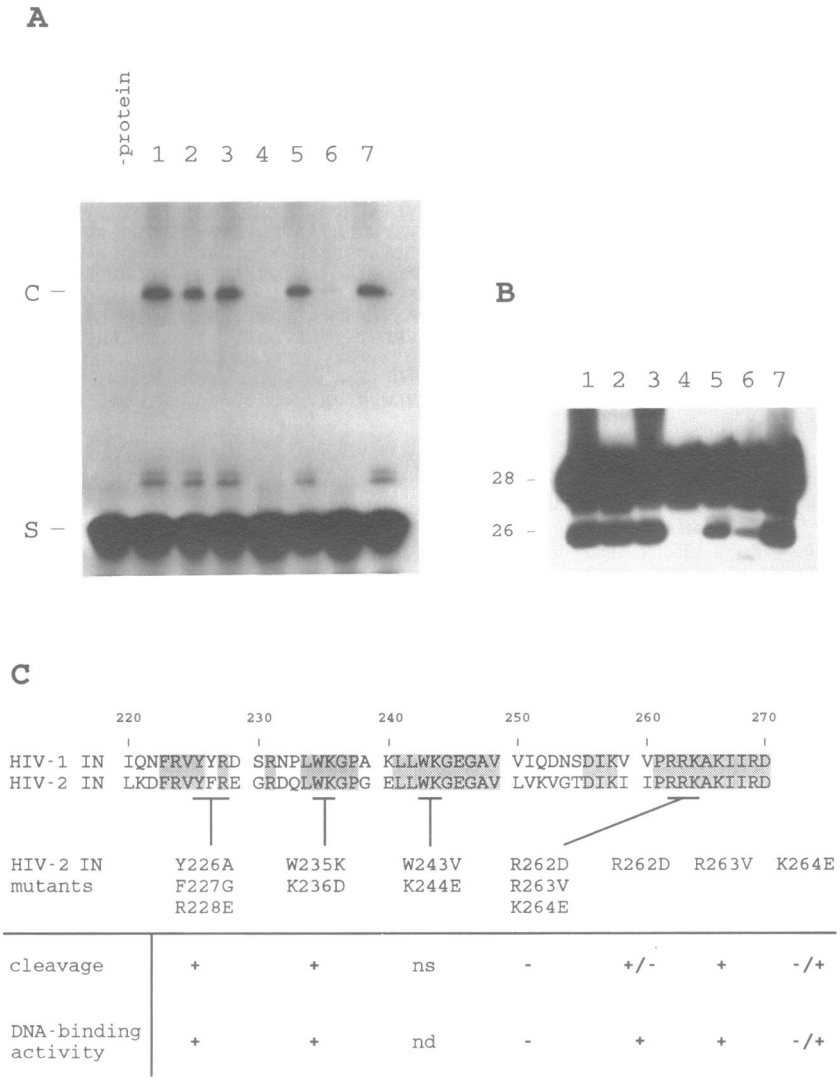

Figure 5. Mutational analysis of the DNA-binding domain of HIV-2 IN. (**A**) DNA-binding activities of polypeptides with point mutations in the DNA-binding domain of HIV-2 IN (MBP−IN-$2_{210-270}$). DNA binding activity was determined by UV crosslinking. −protein; DNA substrate alone; WT (lane 1); Y226A, F227G, R228E (lane 2); W235K, K236D (lane 3);R262D, R263V, K264E (lane 4); R262D (lane 5); K264E (lane 6); R263V (lane 7). UV-crosslinking products were analyzed on a 10% PAGE (S: DNA substrate; C: protein−DNA complex; the nature of the low migrating UV-crosslink products is not understood). (**B**) Site-specific cleavage of oligonucleotides, containing viral U5 LTR sequences by mutant HIV-2 proteins. WT (lane 1), Y226A, F227G, R228E (lane 2), W235K, K236D (lane 3), R262D, R263V, K264E (lane 4), R262D (lane 5), K264E (lane 6), R263V (lane 7). Reaction products were separated on a denaturing 12% polyacrylamide gel and visualized by autoradiography. (**C**) Comparison of amino acid sequences of the DNA-binding domain of HIV-1 and HIV-2. Identical amino acid residues are indicated by shaded boxes. DNA-binding and site-specific cleavage activities of mutant proteins are shown. Activities were determined by densitometry and scored as follows: + >50% of wt activity; +/− 10−50% wt activity; −/+ 1−10% wt activity; − <1% wt activity. ns = not soluble; nd = not determined.

which is not detected in the assays used. One possibility is that IN contains an as yet undefined DNA-binding domain which binds specifically to viral DNA ends and a second non-specific DNA-binding domain in the C-terminus. Another possibility is, that the C-terminal DNA-binding domain can bind DNA in a non-specific manner, but also in a specific manner upon interaction with other parts (domains) of IN, either of the same or another IN monomer. We favour the latter possibility. It was found that IN forms a stable complex with viral DNA in the presence of Mn^{2+}. This stable binding requires the presence of an intact N-terminal domain, active site and C-terminal DNA-binding domain as well as specific viral DNA sequences (37). It remains to be determined how IN specifically recognizes viral LTR sequences. In this respect, biochemical analysis of the viral pre-integration complex will certainly help to understand specific IN−DNA interactions. Several proteins have been detected in this complex (38,39), but it is not yet clear whether all or any of these play an essential role in the integration reaction (e.g. see Lapadat-Tapolsky *et al.* (40)).

In addition, an important contribution to elucidation of the mechanism of retroviral integration is expected from the three dimensional structure of integrase. Until now, attempts to crystallize full-length IN for X-ray analysis have failed because of poor solubility of the protein. Alternatively, structural analysis by X-ray crystallography or NMR analysis of domains of IN (14,15) could be considered. Overproduction and elucidation of

the three dimensional structure of the non-specific DNA-binding domain ($IN_{220-270}$) of IN is now feasible.

ACKNOWLEDGEMENTS

We thank Toby Gibson (EMBL, Heidelberg) for the protein sequence alignment and Chris Vos, Fusinita van den Ent and Piet Borst for critical reading of the manuscript and fruitful discussions. This work is supported by the Dutch Organisation for Scientific Research (NWO; grant 900-502-140).

REFERENCES

1. Stevenson,M., Haggerty,S., Lamonica,C.A., Meier,C.M., Welch,S.K. and Wasiak,A.J. (1990) J. Virol., **64**, 2421–2425.
2. LaFemina,R.L., Schneider,H.L., Robbins,H.L., Callahan,P.L., LeGrow,K., Roth,E., Schleif,W.A. and Emini,E.A. (1992) J. Virol., **66**, 7414–7419.
3. Vink,C. and Plasterk,R.H.A. (1993) Trends in Genet., **9**, 433–437.
4. Goff,S.P. (1992) Annu. Rev. Genet., **26**, 527–544.
5. Skalka,A.M. (1993) Gene, **135**, 175–182.
6. Sherman,P.A. and Fyfe,J.A. (1990) Proc. Natl. Acad. Sci. USA, **87**, 5119–5123.
7. Vink,C., van Gent,D.C., Elgersma,Y. and Plasterk,R.H.A. (1991) J. Virol., **65**, 4636–4644.
8. Bushman,F.D. and Craigie,R. (1991) Proc. Natl. Acad. Sci. USA, **88**, 1339–1343.
9. Leavitt,A.D., Rose,R.B. and Varmus,H.E. (1992) J. Virol., **66**, 2359–2368.
10. van Gent,D.C., Elgersma,Y., Bolk,M.W.J., Vink,C. and Plasterk,R.H.A. (1991) Nucleic Acids Res., **19**, 3821–3827.
11. Drelich,M., Wilhelm,R. and Mous,J. (1992) Virology, **188**, 459–468.
12. Vincent,K.A., Ellison,V., Chow,S.A. and Brown,P.O. (1993) J. Virol., **67**, 425–437. 13. Chow,S.A., Vincent,K.A., Ellison,V. and Brown,P.O. (1992) Science, **255**, 723–726.14. Bushman,F.D., Engelman,A., Palmer,I., Wingfield,P. and Craigie,R. (1993) Proc. Natl. Acad. Sci. USA, **90**, 3428–3432.
15. Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) Nucleic Acids Res., **21**, 1419–1425.
16. Khan,E., Mack,J.P.G., Katz,R.A., Kulkosky,J. and Skalka,A.M. (1991) Nucleic Acids Res., **19**, 851–860.
17. Fayet,O., Ramond,P., Polard,P., Prere,M.F. and Chandler,M. (1990) Mol. Microbiol., **4**, 1771–1777.
18. Burke,C.J., Sanyal,G., Bruner,M.W., Ryan,J.A., LaFemina,R.L., Robbins,H.L., Zeft,A.S., Middaugh,C.R. and Cordingley,M.G. (1992) J. Biol. Chem., **267**, 9639–6944.
19. Engelman,A. and Craigie,R. (1992) J. Virol., **66**, 6361–6369.
20. Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G. and Skalka,A.M. (1992) Mol. Cell. Biol., **12**, 2331–2338.
21. van Gent,D.C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1992) Proc. Natl. Acad. Sci. USA, **89**, 9598–9602.
22. Mumm,S.R. and Grandgenett,D.P. (1991) J. Virol., **65**, 1160–1167.
23. Woerner,A.M., Klutch,M., Levin,J.G. and Marcus-Sekura,C.J. (1992) AIDS Res. Hum. Retr., **8**, 297–304.
24. LaFemina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) J. Virol., **65**, 5624–5630.
25. Schauer,M. and Billich,A. (1992) Biochem. Biophys. Res. Comm., **186**, 874–888.
26. Jones,K.S., Coleman,J., Merkel,G.W., Laue,T.M. and Skalka,A.M. (1994) J. Biol. Chem., **287**, 16037–16040.
27. van Gent,D.C., Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) EMBO J., **12**, 3261–3267.
28. Engelman,A., Bushman,F.D. and Craigie,R. (1993) EMBO J., **12**, 3269–3275.
29. Rosenberg,A.H., Lade,B.N., Chui,D., Lin,S.-W., Dunn,J.J. and Studier,F.W. (1987) Gene, **56**, 125–135.
30. Schaegger,H. and von Jagow,G. (1987) Anal. Biochem., **166**, 368–379.
31. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular cloning: a laboratory manual. Cold spring harbor laboratory press, New York.
32. Woerner,A.M. and Marcus-Sekura,C.J. (1993) Nucleic. Acids Res., **21**, 3507–3511.
33. Leavitt,A.D., Shiue,L. and Varmus,H.E. (1993) J. Biol. Chem., **268**, 2113–2119.
34. Beese,L.S. and Steitz,T.A. (1991) EMBO J., **10**, 25–33.
35. Sherman,P.A., Dickson,M.L. and Fyfe,J.A. (1992) J. Virol., **66**, 3593–3601.
36. Vink,C., van der Linden,K.H. and Plasterk,R.H.A. (1994) J. Virol., **68**, 1468–1474.
37. Vink,C., Puras Lutzke,R.A. and Plasterk,R.H.A. (1994) Nucleic Acids Res., in press.
38. Farnet,C. and Haseltine W.A., (1991) J. Virol., **65**, 1910–1915.
39. Bukrinsky,M.I., Sharova,N., McDonald,T.L., Pushkarskaya,T., Gary Tarpley,W., Stevenson,M. (1993) Proc. Natl. Acad. Sci. USA, **90**, 6125–6129.
40. Lapadat-Tapolsky,M., De Rocquigny,H., van Gent,D., Roques,B., Plasterk,R.H.A. and Darlix,J.-L. (1993), Nucleic Acids Res., **21**, 831–839.
41. Lightfoote,M.M., Coligan,J.E., Folks,T.M., Fauci,A.S., Martin,M.A. and Venkatesan,S. (1986) J. Virol., **60**, 771–775.
42. Adachi,A., Gendelman,H.E., Koenig,S., Folks,T., Willey,R., Rabson,A. and Martin,M.A. (1986) J. Virol., **59**, 284–291.
43. Guyader,M., Emerman,M., Sonigo,P., Clavel,F., Montagner,L. and Alizon,M. (1987) Nature, **326**, 662–669.
44. Talbott,R.L., Sparger,E.E., Lovelace,K.M., Fitch,F.M., Pedersen,N.C., Luciw,P.A. and Elder,J.H (1989) Proc. Natl. Acad. Sci. USA, **86**, 5743–5747.
45. Fomsgaard,A., Hirsch,V.M., Allan,J.S. and Johnson,P.R. (1991) Virology, **182**, 397–402.
46. Garvey,K.L., Oberste,M.S, Elser,J.E., Braun,M.J. and Gonda,M.A. (1990) Virology, **175**, 391–409.
47. Stephens,R.M., Casey,J.W and Rice,N.R. (1986) Science, **231**, 589–594.
48. Whetter,L., Archambault,D., Perry,S., Gazit,A., Coggins,L., Yaniv,A., Clabough,D., Dahlberg,J. and Tronick,S. (1990) J. Virol., **64**, 5750–5756.
49. Anderson,O.S., Elser,J.E., Tobin,G.J., Greenwood,J.D., Gonda,M.A., Georgson,G., Andresdottir,V., Benediktsdottir,E., Carlsdottir,H.M., Mantyla,E.O., Rafnar,B., Palson,P.A., Casey,J.W. and Petursson,G. (1993) Virology, **193**, 89–105.
50. Sonigo,P., Barker,C., Hunter,E. and Wain-Hobson,S. (1986) Cell, **45**, 375–385.
51. Schwartz,R.M., Tizard,R. and Gilbert,W. (1983) Cell, **32**, 853–869
52. Moore,R., Dixon,M., Smith,R., Peters,G. and Dickson,C. (1987) J. Virol., **61**, 480–490.
53. Malik,K.T.A., Even,J. and Karpas,A. (1988) J. Gen. Virol., **69**, 1695–1710.
54. Shimotohno,K., Takahashi,Y., Shimizu,N., Gojobori,T., Golde,D.W., Chen,S.Y., Miwa,M. and Sugimura,T. (1986) Proc. Natl. Acad. Sci. USA, **82**, 3101–3105.
55. Rice,N.R., Stephens,R.M., Burney,A. and Gilden,R.V. (1985) Virology,
56. Shinnik,T.M., Lerner,R.A. and Sutcliffe,J.G. (1981) Nature, **293**, 543–548.